第 39 卷第 8 期
录用定稿
计算机应用研究
Application Research of Computers
Vol. 39 No. 8
Accepted Paper

多尺度特征深度复用的显著性目标检测算法 *

周之平，樊　斌，盖　杉，徐温程

(南昌航空大学 信息工程学院，南昌 330063)

摘　要：针对传统显著性目标检测方法在检测不同尺度的多个显著性目标方面的不足，提出了一种多尺度特征深度复用的显著性目标检测算法。网络模型由垂直堆叠的双向密集特征聚合模块和水平堆叠的多分辨率语义互补模块组成。首先，双向密集特征聚合模块基于 ResNet 骨干网络提取不同分辨率语义特征，然后，依次在 top-down 和 bottom-up 两条通路上进行自适应融合，以获取不同层次多尺度表征特征；最后，通过多分辨率语义互补模块对两个相邻层次的多尺度特征进行融合，以消除不同层次上特征之间的相互串扰，来增强预测结果的一致性。在 5 个基准数据集上进行的实验结果表明，该方法在 F_{max}、S_m、MAE 最高能达到 0.939、0.921、0.028，且检测速率可达 74.6 fps，与其他对比算法相比有着更好的检测性能。

关键词：显著性目标检测；多尺度特征；双向密集特征聚合；多分辨率语义；深度学习

中图分类号：TP389.1　　doi: 10.19734/j.issn.1001-3695.2022.01.0033

Deep multiplexing multi-scale features for salient object detection

Zhou Zhiping, Fan Bin, Gai Shan, Xu Wencheng

(*School of Information Engineering*, *Nanchang Hangkong University*, *Nanchang* 330063, *China*)

Abstract: In view of the shortcomings of traditional salient target detection methods in detecting multiple salient targets at different scales. This paper presents a salient object detection algorithm with deep multiplexing of multi-scale features. The network model consists of vertically stacked bidirectional dense feature aggregation modules and horizontally stacked multi-resolution semantic complementary modules. First, the bidirectional dense feature aggregation module extracts semantic features of different resolutions based on the ResNet backbone network, and then performs adaptive fusion on the top-down and bottom-up paths in turn to obtain multi-scale representation features at different levels; The multi-resolution semantic complementation module fuses the multi-scale features of two adjacent levels to eliminate the mutual crosstalk between features at different levels and enhance the consistency of prediction results. The experimental results on 5 benchmark datasets show that the method can achieve the highest F_{max}, S_m, MAE of 0.939, 0.921, 0.028, and the detection rate can reach 74.6 fps, which has better detection performance compared with other comparison algorithms .

Key words: salient object detection; multi-scale features; bidirectional dense feature aggregation; multi-resolution semantic; deep learning

0　引言

显著性目标检测(SOD： Salient Object Detection)是计算机视觉领域的一项关键技术，旨在从输入图像中分割出视觉上最为明显的区域。深度学习技术的兴起促进了 SOD 技术的大力发展，并将 SOD 性能提升到一个新的水平。SOD 已被广泛应用于计算机视觉的多个领域，如图像分割[1]、视觉跟踪[2]、图像质量评估[3]、图像检索[4]、边缘检测[5]等。在基于 CNN 的 SOD 模型中，不同层次的特征可以表征显著性对象的不同特性。具体来说，低层语义特征有着显著对象的详细信息，但包含大量噪声，而高级语义特征可帮助网络定位显著对象的位置，但缺乏有关对象的详细信息。

如何从尺度变化的数据中提取更有效的信息，以及怎么使得预测结果与图像中的显著目标在空间上保持一致，仍然是两个悬而未决的问题。近年来所开展的研究工作专注于设计复杂的网络结构，提取具有强辨识能力的多尺度特征或对多尺度特征进行高效地融合，以满足对不同尺度显著性目标检测的要求。Zhang 等人[6]提出了一个通用的聚合多级卷积特征的框架，它以完全连接的方式组合来自多层的特征。Hou 等人[7]在整体嵌套边缘检测模型(HED)中引入快捷连接，提出了一种跳层结构，该结构具有一系列从高级特征到低级特征的快捷连接。Liu 等人[8]通过选择性地聚合上下文信息来构建全局上下文特征，然后将全局上下文和多尺度局部上下文进行归并来提升效果。Wu 等人[9]提出了一种新颖的级联部分解码器框架，该框架丢弃了低层特征以降低深度聚合模型的复杂性，并利用生成的相对精确的注意力图来精炼高层特征。Pang 等人[10]提出聚合交互模块，通过相互学习的方式有效地利用相邻层的特征和自适应模块，使网络自适应地提取多尺度信息，以更好地处理尺度的变化。

此外，为了生成更好的显著性映射图，有必要对多层次的特征进行归并。然而，过度集成不同分辨率的特征，不仅会带来大量的计算开销，还会导致有用特征被稀释，进而使算法性能退化。为此，研究学者提出了多种不同的方案以克服这一问题。Feng 等人[11]采用每个编码器块和相应的解码器块构建的注意反馈模块来帮助结合多层次特征。Wei 等人[12]采用选择性融合策略，通过元素级乘法操作来融合不同层次的特征以抑制冗余信息，避免不同层次特征之间的相互污染。Qin 等人[13]提出了一个两层嵌套的 U 型结构来集成多层次的

收稿日期：2022-01-04；**修回日期**：2022-03-11　**基金项目**：国家基金资助项目(62061032)

作者简介：周之平(1975-)，男，江西南昌人，讲师，硕导，博士，主要研究方向为目标检测(zhouzhp@126.com)；樊斌(1997-)，男，江西九江人，硕士研究生，主要研究方向为深度学习、目标检测等；盖杉(1980-)，男，吉林长春人，教授，硕导，博士，主要研究方向为模式识别等；徐温程(1997-)，男，河北廊坊人，硕士研究生，主要研究方向为目标检测等.

深层特征。Chen 等人[14]将残差学习引入到 HED 的体系结构中，在自上而下的路径中使用反向注意力来指导残余显著性学习，引导网络能快速而有效地发现缺失的对象部分和缺损细节。陈等人[15]提出结合中心邻域对比度机制和卷积神经网络，为有效提高多尺度特征的表示能力提供了一种强有力的方法。

综上所述，如何将 CNN 主干网络中不同层次的特征有效融合至关重要。该文基于 UNet 网络模型[16]，提出一种多尺度特征深度复用的显著性目标检测模型(DMMF:deep multiplexing multi-scale feature)。该模型中设计了双向密集特征聚合模块(BDA: bidirectional dense aggregation module)，在 top-down 和 bottom-up 两条通路中对主干网抽取的不同分辨率的 CNN 特征进行重用，并利用残差连接进行特征增强。通过堆叠多个不同尺度的 BDA 模块来提取具有多种分辨率语义的多层次特征。结合文献[10]的设计理念，设计了多分辨率语义互补模块(MSC: Multi-resolution semantic complement module)，按照级联方式将其植入到 UNet 网络的 bottom-up 通路中，以增强模型对显著性目标的预测能力。

1　多尺度特征深度复用网络

多尺度特征深度复用的显著性目标检测网络如图 1 所示。该网络以 Resnet50 作为主干网络，为了使初始特征更加多样化，提出了一个堆叠的双向密集特征聚合模块对主干网络提取的特征进行全分辨率融合，提取语义更为丰富的多尺度特征。对于获取到的多个层次的多尺度特征，采用级联的多分辨率语义互补模块来保留相邻特征节点中的有用信息，逐级还原显著性目标的语义信息和空间信息。

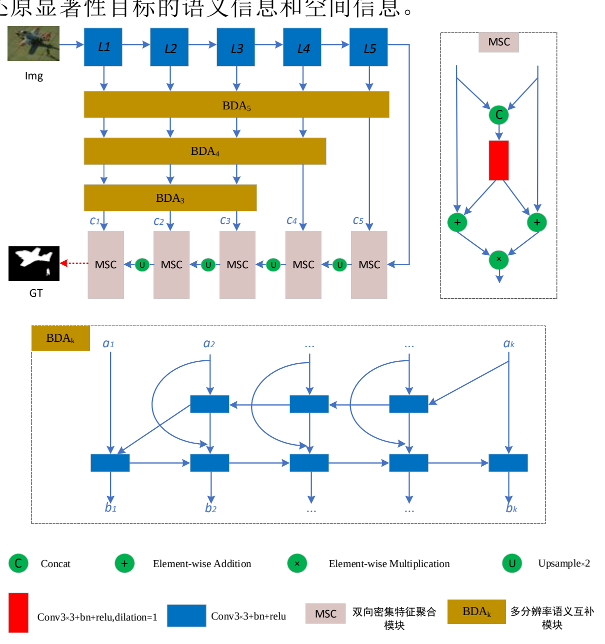


图 1　DMMF 网络框架图

Fig. 1　DMMF network framework diagram

1.1　双向密集特征聚合模块

双向密集特征聚合模块旨在聚合不同分辨率的特征。形式上，给出多尺度列表特征 $L_{in}=(L_{in}^{l_1}, L_{in}^{l_2},\ldots)$，其中 $L_{in}^{l_i}$ 表示 l^i 层的特征，该算法的目标是找到一个可以有效聚合不同特征并输出新特征列表的变换 $f: L_{out}=f(L_{in})$。传统的 FPN[17]以自顶向下的方式聚合多尺度特征,其本质上受到单向信息流的限制。为了解决这个问题，PANet[18]增加了一个额外的自下而上的路径聚合网络，性能提升的同时带来了更多的参数和计算。NAS-FPN[19]使用神经架构搜索来搜索更好的跨尺度特征网络拓扑，但在搜索过程中需要数千个 GPU 小时，并且发现的网络不规则且难以解释或修改。EfficientDet[20]通过对 PANet 进行化简构建了 BiFPN 模块，并通过多次堆叠 BiFPN 以更有效地获得更具鉴别性的多尺度特征。

双向密集特征聚合模块通过双向(自顶向下和自底向上)跨尺度连接路径进行多尺度特征提取，当融合不同分辨率的特征时，由于输入特征具有不同的分辨率，它们通常对输出特征的贡献不均等,该算法通过一个简单的注意力机制为每个输入增加一个额外的权重，让网络学习每个输入特征的重要性。然而，不同于简单的级联操作，该算法通过堆叠不断减小规模的双向密集特征聚合模块来实现更高级别的特征融合，以更少量的参数达到相同甚至更优的效果。

BDA 模块的结构如图 1 所示。以图 2 的 BDA5 为例，下标“5”表示该模块有 5 个输入信号，对应于 Resnet50 网络 5 个 stage 上提取到的基本特征 $a_1\sim a_5$。首先，a_5 通过上采样与 a_4 将进行逐元素线性加权操作以及 3×3 卷积(含批归一化层、Relu 激活函数层)操作获得 m_4，同理自底向上依次获得 m_3，m_2 以及 b_1。然后，b_1 下采样与 m_2、a_2 进行逐元素线性加权操作以及 3×3 卷积(含批归一化层、Relu 激活函数层)获得 b_2，同理自顶向下的依次获得 b_3，b_4，b_5。最后，将 b_5 作为 MSC 的输入之一 c_5，同时 b_1，b_2，b_3，b_4 作为模块 BDA4 的输入。与 BDA5 类似，BDA4 将得到 b_1，b_2，b_3，b_4 四个输出，b_4 将作为 MSC 的输入之一 c_4，同时 b_1，b_2，b_3 作为模块 BDA3 的输入，最终 BDA3 的三个输出作为 MSC 的输入 c_1，c_2，c_3。

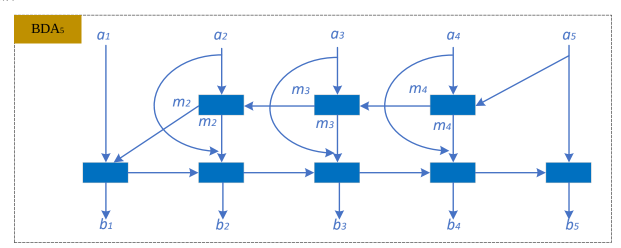


图 2　双向密集特征聚合模块

Fig. 2　Bidirectional dense aggregation module

图 2 中的 BDA5 的融合过程如式(1)和(2)所示。

$$m_k=\begin{cases}0 & k=1,5\\ conv(w_1\times a_k+w_2\times up(a_{k+1})) & 1<k<5\end{cases} \tag{1}$$

$$b_k=\begin{cases}conv(w_1\times a_k+w_2\times up(m_{k+1})) & k=1\\ conv(w_1\times a_k+w_2\times m_k+w_3\times down(b_{k-1})) & 1<k<5\\ conv(w_1\times a_k+w_3\times down(b_{k-1})) & k=5\end{cases} \tag{2}$$

其中: m_k，k=2~K 是自上而下路径的中间特征，conv 为对特征进行 3×3 卷积，再加上批归一化(Batch Normalization)和 Relu 激活函数的一组操作，w_i 为特征融合阶段给每个输入所分配的权重系数，w_i 初始化为(0,1)的随机数，并利用 Laplace 平滑进行归一化处理: $w_i:=(w+\varepsilon)/\sum_i(w+\varepsilon), i=1,2,3$，其中 $\varepsilon=1\times10^{-4}$，用于避免数值计算的不稳定。网络每次训练后更新 w_i，使用 Relu 函数保证其非负，并通过 Laplace 平滑重新归一化处理。

1.2　多分辨率语义互补模块

堆叠的双向密集特征聚合模块用于从骨干网络提取不同层次的有效的多尺度特征，而多分辨率语义互补模块则是让相邻层次的多尺度特征在空间、语义上相互补充，不断增强适合当前分辨率的特征，削弱不合适的特征，进而找到适合当前输入信息的特征。

自上而下经过 BDA5~BDA3 获得了一组最终地多尺度语义特征 c_1，c_2，c_3，c_4，c_5 分辨率依次为 320×320，160×160，80×80，40×40，20×20，这些特征都含有来自不同尺度目标的语义信息，但各个语义成分的重要程度存在差异。如果简单地将这些特征进行线性融合会造成特征之间的相互干扰，弱化那些具有较强鉴别能力的特征，进而影响算法的检测性能。为此，提出多分辨率语义互补模块，以充

分挖掘各个粒度特征中的有用信息，形成优势互补。MSC 的详细情况如图 3。

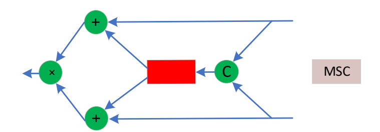


图 3　多分辨率语义互补模块

Fig. 3　Multi-resolution semantic complement module

MSC 的详可以表示为

$$f_g = conv(concat(f_1, f_2)) \quad (3)$$

$$f_1^* = f_1 \oplus f_g \quad (4)$$

$$f_2^* = f_2 \oplus f_g \quad (5)$$

$$f_c = conv(f_1^*, f_2^*) \quad (6)$$

其中，f_1，f_2 表示相邻特征，conv 表示带批量归一化层和 ReLU 激活函数的空洞卷积，$\oplus$ 为逐元素加法，$\otimes$ 为逐元素乘法。MSC 首先通过 concat 操作将输入特征进行合并，然后通过 rate 为 1 的空洞卷积，同时，进行归一化和 relu 操作，这样就获取到了融合了两个输入特征的全局语义信息。然后通过逐元素相加将得到的全局语义信息添加回输入特征，来分别对输入特征进行空间和语义上的补强。最后通过逐元素乘法，并加入自适应权重，来有选择的继承两组从空间和语义上得到补强的特征。这样，MSC 就实现了让输入特征继承重要特性，同时丢弃更多噪声的目标。

2　实验和结果

2.1　数据集

训练数据集：新方法在 DUTS-TR 上进行网络训练，DUTS-TR 是 DUTS 数据集的一个子集，总共包含 10553 幅图像。它是目前规模最大和最常用的显著性目标检测训练数据集。为了确保模型的收敛，设置训练的轮次为 80，采用 SGD 优化器，初始学习率为 1×10^{-3}，权重衰减为 5×10^{-4}，其中，动量项系数为 0.9。所有实验均在 Linux 16.04 操作系统，GPU(GTX TITAN-XP)，Pytorch 1.0.0，cuda9.0 环境下完成。

测试数据集：使用 6 个常用的基准数据集对提出的方法进行评估，包括：DUT-OMRON[21]，DUTS-TE[22]，HKU-IS[23]，ECSSD[24]，pascal-S[25]。DUT-OMRON 包括 5168 幅图像，其中大多数包含一个或两个结构复杂的前景目标。DUTS 数据集由 DUTS-TR 和 DUTS-TE 两部分组成。因为算法训练时使用了 DUTS-TR 数据集，为此选择包含 5019 幅图像的 DUTS-TE 进行测试。HKU-IS 包含 4447 幅图像，其中包含多个与图像边界相交的不连续显著对象。ECSSD 包含 1000 幅结构复杂的图像，多数图像包含尺度较大的前景目标。PASCAL-S 包含 850 幅图片，这些图像都有着复杂的前景对象和杂乱的背景。

2.2　评估指标

为了更全面地进行实验评价，该算法选择了 F-measure，平均绝对精度(MAE)，S-measure 三个广泛使用的评价指标来对算法的性能进行评价。

F-measure 是综合 Precision 和 Recall 的加权调和平均值，定义如下：

$$F_\beta = \frac{(1+\beta^2)\times Precision \times Recall}{\beta^2 \times Precision \times Recall} \quad (7)$$

其中，β^2 一般设置为 0.3，F-measure 越大表示预测结果越准确。该算法选择所有阈值计算出的最大值作为评价结果。

平均绝对误差(MAE)是计算预测的显著图与真值图之间的平均绝对误差，计算公式如下所示。

$$MAE = \frac{1}{H\times W}\sum_{x=1}^{H}\sum_{y=1}^{W}|P(x,y)-G(x,y)| \quad (8)$$

其中，P 代表预测的显著图，G 代表对应的真值图。(H,W)代表图像的大小，(x,y)代表像素点的对应坐标。MAE 越小表示预测结果越好。

S-measure 是通过度量目标感知(object-aware) S_o 和区域感知(region-aware) S_r 的结构相似性来评估预测的显著图和真值图之间的结构相似性的评价指标，计算公式如下所示。

$$S = \alpha \times S_o + (1-\alpha)S_j \quad (9)$$

其中，α 通常设置为 0.5, S-measure 越大，表示检测的显著图与真值图在空间结构上越相似。

2.3　性能分析

在 DUTS-TE、DUT-OMRON、HKU-IS、ECSSD 和 PASCAL5 个显著性检测数据集上，使用上述评价指标，将新提出的方法与目前最先进的 11 种方法进行比较，结果如表 1 所示。

表 1　不同算法在测试集上 $F_{\max}$，S_m，MAE 指标的对比

Tab. 1　Comparison of $F_{\max}$, S_m, MAE indicators of different algorithms on the test set

Approachs	DUTS-TE			DUT-OMRON			HKU-IS			ECSSD			PASCAL_S		
	Fmax↑	Sm↑	MAE↓	Fmax↑	Sm↑	MAE↓	Fmax↑	Sm↑	MAE↓	Fmax↑	Sm↑	MAE↓	Fmax↑	Sm↑	MAE↓
MWS[26]	0.789	0.792	0.106	0.718	0.751	0.114	0.841	0.820	0.072	0.878	0.866	0.096	0.790	0.746	0.134
RAS[14]	0.831	0.839	0.059	0.787	0.814	0.061	0.913	0.887	0.045	0.921	0.893	0.056	0.838	0.795	0.104
R3Net[27]	0.833	0.836	0.057	0.795	0.817	0.062	0.915	0.895	0.035	0.934	0.910	0.040	0.846	0.805	0.094
CPD[9]	0.865	0.869	0.046	0.797	0.825	0.056	0.925	0.906	0.034	0.939	0.918	0.037	0.872	0.847	0.072
AFNet[11]	0.863	0.867	0.043	0.801	0.826	0.057	0.925	0.905	0.036	0.935	0.914	0.042	0.871	0.850	0.071
PoolNet[28]	0.880	0.871	0.040	0.808	0.836	0.056	0.932	0.917	0.033	0.944	0.921	0.039	0.865	0.832	0.075
BASNet[30]	0.872	0.879	0.040	0.803	0.832	0.059	0.930	0.913	0.032	0.945	0.924	0.035	0.879	0.853	0.070
EGNet[29]	0.890	0.887	0.039	0.815	0.841	0.053	0.935	0.918	0.031	0.947	0.925	0.037	0.878	0.853	0.075
U2Net[13]	0.873	0.861	0.044	0.823	0.847	0.054	0.935	0.916	0.031	0.951	0.928	0.033	0.859	0.844	0.074
F3Net[12]	0.891	0.888	0.035	0.813	0.838	0.053	0.937	0.917	0.028	0.945	0.924	0.033	0.872	0.855	0.062
MINet[10]	0.884	0.884	0.037	0.811	0.833	0.055	0.935	0.920	0.028	0.947	0.925	0.033	0.882	0.857	0.064
DMMF	0.891	0.890	0.034	0.817	0.845	0.052	0.939	0.921	0.028	0.949	0.930	0.031	0.879	0.856	0.065

由表 1 可知：提出的方法有着很好的性能，在大多数数据集的比较指标上，表现优于其他显著性检测模型。其中，在 HKU-IS 数据集上，该算法表现最好，在 3 个评价指标上都优于其他方法，其中 $F_{\max}$ 比 U2Net 提高了 0.004，S_m 比 U2Net 提高了 0.005,MAE 比 U2Net 减小了 0.003；在 DUTS-TE 数据集上，提出的方法的 $F_{\max}$ 和 S_m 均优于其他方法，只有 MAE 略低于 F3Net；在 ECSSD 数据集上，提出的方法的 MAE 和 S_m 均优于其他方法，只有 $F_{\max}$ 略低于 U2Net。因此，综合所有的数据集和评价指标，提出的方法对多个显著目标和尺度变化有着很好的性能体现。此外，在数据集 ECSSD 上不同方法之间的平均速度(FPS)比

较如表 2 所示。

表 2 不同算法的检测速度指标 FPS 比较

Tab. 2 Comparison of detection speed index FPS of different algorithms

	本文	MWS	RAS	R3Net	CPD	AFNet
Size	320×320	256×256	256×256	224×224	352×352	224×224
FPS	74.6	52	45	33	66	26
	PoolNet	BASNet	EGNet	U2Net	F3Net	MINet
Size	300×400	256×256	256×256	320×320	352×352	320×320
FPS	30	25	25	30	46	86

从表 2 可以看出，提出的算法的检测速度 74.6FPS，仅次于 MINet 的 86FPS，但检测性能优于 MINet。

为了更直观的说明提出的算法的优势，将 11 种最先进的检测方法在不同场景下的预测结果可视化，比较结果如图 4 所示。

图 4 中，第 **1** 行是小的显著性目标的情况，第 **2** 行是大的显著性目标的情况，第 **3,4** 行是包含多个大小不同的显著性目标的情况，第 **5** 行是前景、背景对比度较低的情况，第 **6** 行是在复杂场景下的情况。

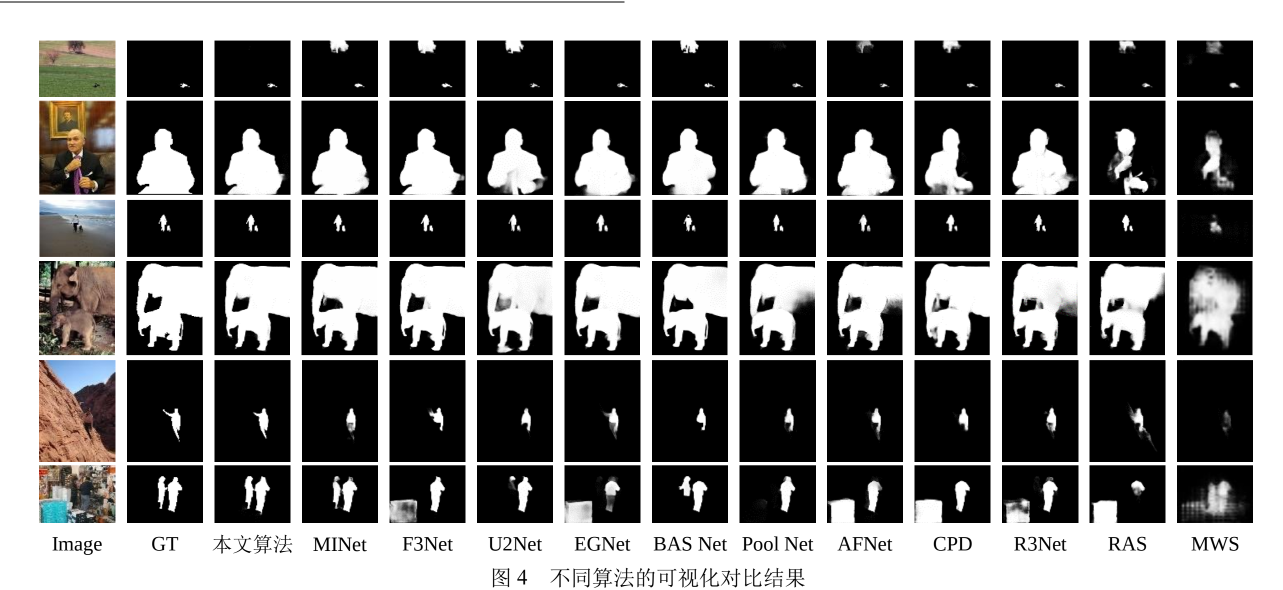


图 4 不同算法的可视化对比结果

Fig. 4 Visual comparison results of different algorithms

由图 4 可知，提出的算法在小目标，大目标，具有不同尺度目标，复杂背景等情况下的检测效果和真值图都相差无几。此外，较其他算法，在小目标情况下，可以有效屏蔽背景干扰，检测到小的显著目标；在大目标情况下能更完整的检测除显著性目标，不会出现缺损。在具有不同尺度目标时也能很好的检测到物体的轮廓边缘。在前景、背景对比度低的情况下，能很好的找到目标的完整轮廓。在复杂背景场景下，也能有效的检测到显著目标而不会检测到干扰目标。

2.4 消融分析

a) 不同 BDA 模块对算法性能的影响

为了获得更好的多尺度信息，该算法巧妙地堆叠多个 BDA 模块。为了验证如何堆叠更利于模型预测，在 HKU-IS 数据集上，对不同堆叠方案进行了测试和比较，结果如表 3 所示，其中 “ $\times n$ ” 表示对结构相同的重复堆叠 n 次。

表 3 不同堆叠方式对算法性能的影响

Tab. 3 Performance comparison of different stacking methods

堆叠方式	F_{max}	MAE
BDA5	0.922	0.036
BDA5×2	0.928	0.033
BDA5×3	0.933	0.031
BDA5×4	0.926	0.034
BDA5+BDA4+BDA3	0.939	0.028

从表 2 数据可以看出，重复对 BDA5 模块堆叠多次时算法性能比使用单个模块要好，这说明多次对多分辨率特征进行融合有助于提升算法性能。但当堆叠次数超过 3 时，算法性能会下降，因为随着网络模型变深，容易导致梯度消失，从而使得网络更难以训练。而对 3 个不同构型的 BDA 模块进行堆叠时，算法性能最优，这反映了对不同分辨率的多层语义特征融合，能防止有用的特征被稀释，从而捕获更具判别性的抽象特征。

b) MSC 和 BDA 组合策略对算法性能的影响

为了验证 MSC 和 BDA 两个模块在模型中的有效性，针对不同组合策略，在 DUTS-TE 数据集上进行测试。使用 F-measure、MAE 和 S-measure 3 个评价指标来进行性能比较。结果如表 4 所示。

表 4 MSFF 算法的消融实验

Tab. 4 Ablation experiment of MSFF algorithm

Baseline	BDAs	BDAs*	MSC	F_{max}	MAE	S_m
√				0.853	0.055	0.842
√	√			0.872	0.046	0.875
√			√	0.878	0.044	0.879
√		√	√	0.885	0.038	0.884
√	√		√	0.891	0.035	0.890

其中：BaseLine 为原始的 U-Net 网络模型，BDAs 为按“BDA_5+BDA_4+BDA_3”堆叠的子网络，BDAs*表示不进行加权的 BDAs，即式 1)和 2)中所有权重 w_i 取为 1。从表 3 数据可以看出，在 Baseline 上引入 **BDAs** 或 **MSC** 模块后算法性能都能获得一定程度的提升。而 **Baseline+BDAs+MSC** 策略获得的性能最佳，相较于 Baseline，模型的 F_{max}，S_m 指标分别提高了 0.038 和 0.048，MAE 下降了 0.02。这说明堆叠多个 BDA 和 MSC 模块能抽取更利于检测任务的抽象特征，引入自适应加权策略融合不同层次的特征可避免特征之间的相互干扰，进而使得预测结果与图像中显著性目标更为一致。

3 结束语

为了克服显著性目标检测中多尺度问题，提出一种基于多尺度特征深度复用的显著性目标检测方法。该方法设计了一个双向密集特征聚合模块，对主干网络提取的卷积特征进行多次重用，利用自适应加权融合特征，以消除不同层次特

征的相互干扰；设计了多分辨率语义互补模块模块，对分辨率相邻的两组特征进行融合，在空间和语义上进行相互增强。测试结果表明：提出的方法在的 F_{max} ， S_m ，MAE 分别能达到 0.939，0.921，0.028，均优于 11 种最先进的方法，且能准确地检测到图像中不同尺度的多个目标，有效地处理背景较为复杂的场景。在下一步的研究工作中，将引入多监督的思想和采用新的注意力机制，来更有力的寻找显著目标的轮廓，并使用深度可分离卷积来减少模型的参数。

参考文献：

[1] 李锋林，李亮. 基于显著性检测的目标图像分割算法 [J]. 电子科技, 2017, 30 (1): 69-71. (Li Fenglin, Li Liang. Target Image Segmentation Algorithm Based on Saliency Detection [J]. Electronic Science and Technology, 2017, 30 (1): 69-71)

[2] Wang Yong, Wei Xian, Lu Ding, *et al.* A robust visual tracking method via local feature extraction and saliency detection [J]. The Visual Computer, 2020, 36 (4): 683-700.

[3] 陈晨. 基于视觉感知的图像质量评价方法研究 [D]. 西安：西安电子科技大学, 2019.

[4] Wang Haoxiang, Li Zhihui, Yang Li, *et al.* Visual saliency guided complex image retrieval [J]. Pattern Recognition Letters, 2020, 130 (2020): 64-72.

[5] 张艳邦，张芬，张姣姣. 基于图像边缘特征的目标检测算法 [J]. 内江科技，2021, 42 (04): 47-67. (Zhang Yanbang, Zhang Fen, Zhang Jiaojiao. Target detection algorithm based on image edge features [J]. Neijiang Science and Technology, 2021, 42 (04): 47-67)

[6] Zhang Pingping, Wang Dong, Lu Huchuan, *et al.* Amulet: Aggregating multi-level convolutional features for salient object detection [C]// Proc of the IEEE International Conference on Computer Vision. IEEE: MIT Press, 2017: 202-211.

[7] Hou Qibin, Cheng Mingming, Hu Xiaowei, *et al.* Deeply supervised salient object detection with short connections [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2017: 3203-3212.

[8] Liu Nian, Han Junwei, Yang Ming-Hsuan. Picanet: Learning pixel-wise contextual attention for saliency detection [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2018: 3089-3098.

[9] Wu Zhe, Li Su, Huang Qingming. Cascaded partial decoder for fast and accurate salient object detection [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2019: 3907-3916.

[10] Pang Youwei, Zhao Xiaoqi, Zhang Lihe, *et al.* Multi-scale interactive network for salient object detection [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2020: 9413-9422.

[11] Feng Mengyang, Lu Huchuan, Ding Errui. Attentive feedback network for boundary-aware salient object detection [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2019: 1623-1632.

[12] Wei Jun, Wang Shuhui, Huang Qingming. F³Net: Fusion, Feedback and Focus for Salient Object Detection [C]// Proc of the AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Presss, 2020, : 12321-12328.

[13] Qin Xuebin, Zhang Zichen, Huang Chenyang, *et al.* U2-Net: Going deeper with nested U-structure for salient object detection [J]. Pattern Recognition, 2020, 106: 107404.

[14] Chen Shuhan, Tan X, Wang Ben, *et al.* Reverse attention for salient object detection [C]// Proc of the European Conference on Computer Vision (ECCV) . Berlin: Springer, 2018: 234-250.

[15] 陈琴，朱磊，后云龙等. 基于深度中心邻域金字塔结构的显著目标检测 [J]. 模式识别与人工智能, 2020, 33 (06): 496-506. (Chen Qin, Zhu Lei, Hou Yunlong, etc. Salient object detection based on depth center neighborhood pyramid structure [J]. Pattern Recognition and Artificial Intelligence, 2020, 33 (06): 496-506)

[16] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [C]// International Conference on Medical image computing and computer-assisted intervention. Berlin: Springer, 2015: 234-241.

[17] Lin T Y, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2017: 2117-2125.

[18] Mei Yiqun, Fan Yuchen, Zhang Yulun, *et al.* Pyramid attention networks for image restoration [J]. arXiv preprint arXiv: 2004. 13824, 2020.

[19] Ghiasi G, Lin T Y, Le Q V. Nas-fpn: Learning scalable feature pyramid architecture for object detection [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2019: 7036-7045.

[20] Tan Mingxing, Pang Ruoming, Le Q V. Efficientdet: Scalable and efficient object detection [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2020: 10781-10790.

[21] Yang Chuan, Zhang Lihe, Lu Huchuan, *et al.* Saliency detection via graph-based manifold ranking [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2013: 3166-3173.

[22] Wang Lijun, Lu Huchuan, Wang Yifan, *et al.* Learning to detect salient objects with image-level supervision [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2017: 136-145.

[23] Li Guanbin, Yu Yizhou. Visual saliency based on multiscale deep features [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2015: 5455-5463.

[24] Yan Qiong, Xu Li, Shi Jianping, *et al.* Hierarchical saliency detection [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2013: 1155-1162.

[25] Li Yin, Hou Xiaodi, Koch C, *et al.* The secrets of salient object segmentation [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2014: 280-287.

[26] Zeng Yu, Zhuge Y, Lu Huchuan, *et al.* Multi-source weak supervision for saliency detection [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2019: 6074-6083.

[27] Deng Zijun, Hu Xiaowei, Zhu Lei, *et al.* R3net: Recurrent residual refinement network for saliency detection [C]// Proc of the 27th International Joint Conference on Artificial Intelligence. Palo Alto, CA AAAI Press, 2018: 684-690.

[28] Liu Jiangjiang, Hou Qibin, Cheng Mingming, *et al.* A simple pooling-based design for real-time salient object detection [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2019: 3917-3926.

[29] Zhao Jiaxing, Liu Jiangjiang, Fan Dengping, *et al.* EGNet: Edge guidance network for salient object detection [C]// Proc of the IEEE/CVF International Conference on Computer Vision. 2019: 8779-8788.

[30] Qin Xuebin, Zhang Zichen, Huang Chenyang, *et al.* Basnet: Boundary-aware salient object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE: MIT Press, 2019: 7479-748